

unofficial amendment/for interview discussion only

## ROTHWELL, FIGG, ERNST & MANBECK

1425 K Street, NW  
Suite 800  
Washington, D.C. 20005

Telephone: (202)783-6040  
Telefax: (202)783-6031

### FACSIMILE TRANSMITTAL SHEET

Date: December 6, 2010

To: Examiner Shubo (Joe) Zhou  
U.S. Patent & Trademark Office

Re: U.S. Serial Number 10/539,208  
Our Ref.: 2958-147

Fax. No.: 571 273-0724

From: Monica Chin Kitts

**Please confirm receipt of this telefax**

Number of Pages Including This Transmittal Sheet: 19

If any problems in connection with this facsimile, please contact: Cindy 202-624-1584

*initials*

Draft claims for discussion at interview on 12/8/10

1. (Currently amended) A computerized method for optimizing a nucleotide sequence for the expression of a protein on the basis of the amino acid sequence of the protein, which comprises the following steps carried out on a computer:

generation of generating, using a computer, a first test sequence of n codons which correspond to n consecutive amino acids in the protein sequence, where n is a natural number and is less than or equal to N, the number of amino acids in the protein sequence,

specification of specifying, using a computer, m optimization positions in the test sequence which correspond to the position of m codons at which the occupation by a codon, relative to the test sequence, is to be optimized, where  $m \leq n$  and  $m < N$ ,

generation of generating, using a computer, one or more further test sequences from the first test sequence by replacing at one or more of the m optimization positions a codon of the first test sequence by another codon which expresses the same amino acid,

assessment of assessing, using a computer, each of the test sequences with a quality function and ascertaining the test sequence which is optimal in relation to the quality function,

specification of specifying, using a computer, p codons of the optimal test sequence which are located at one of the m optimization positions, as result codons which form the codons of the

optimized nucleotide sequence at the positions which corresponds to the position of said p codons in the test sequence, where p is a natural number and  $p \leq m$ ,

~~iteration~~ iterating the preceding steps, where in each iteration step the test sequence comprises the appropriate result codon at the positions which correspond to positions of specified result codons in the optimized nucleotide sequence, and the optimization positions are different from positions of result codons.

2. (Original) The method as claimed in claim 1, characterized in that in one or more iteration steps the m optimization positions of the test sequences directly follow one or more result codons which have been specified as part of the optimized nucleotide sequence.

3. (Previously presented) The method as claimed in claim 1, characterized in that in one or more iteration steps the p codons which are specified as result codons of the optimized nucleotide sequence are p consecutive codons.

4. (Previously presented) The method as claimed in claim 1, characterized in that in one iteration step test sequences with all possible codon occupations for the m optimization positions are generated from the first test sequence, and the optimal test sequence is ascertained from these test sequences.

5. (Previously presented) The method as claimed in claim 1,

characterized by:

- assessment of each test sequence with a quality function,
- ascertaining of an extreme value within the values of the quality function for all partial sequences generated in an iteration step,

- specification of p codons of the test sequence which corresponds to the extremal value of the weight function as result codons at the appropriate positions, where p is a natural number and  $p \leq m$ .

6. (Original) The method as claimed in claim 5, characterized in that the quality function takes account of one or more of the following criteria:

- codon usage for a predefined organism, GC content, repetitive sequences, secondary structures, inverse complementary sequence repeats and sequence motifs.

7. (Previously presented) The method as claimed in claim 6, characterized in that the quality function is a function of various single terms which in each case assess one criterion from the following list of criteria:

- codon usage for a predefined organism, GC content, sequence motifs, repetitive sequences, secondary structures, inverse complementary sequence repeats.

8. (Previously presented) The method as claimed in claim 1, characterized in that the quality function takes account of one or more of the following criteria:

- exclusion of inverse complementary sequence identities of more than 20 nucleotides to the transcriptome of a predefined organism,
- exclusion of homology regions of more than 100 base pairs to a predefined DNA sequence,

- exclusion of homology regions with more than 90% similarity of the nucleotide sequence to a predefined DNA sequence.

9. (Previously presented) The method as claimed in claim 1, characterized by the step of synthesizing the optimized nucleotide sequence.

10. (Original) The method as claimed in claim 9, characterized in that the step of synthesizing the optimized nucleotide sequence takes place in a device for automatic synthesis of nucleotide sequences which is controlled by the computer which optimizes the nucleotide sequence.

11. (Previously presented) A device for optimizing a nucleotide sequence for the expression of a protein on the basis of the amino acid sequence of the protein, which has a computer unit comprising algorithms, including:

- an algorithm for generation of a first test sequence of n codons which correspond to n consecutive amino acids in the protein sequence, where n is a natural number and is less than or equal to N, the number of amino acids in the protein sequence,
- an algorithm for specification of m optimization positions in the test sequence which correspond to the position of m codons at which the occupation by a codon, relative to the test sequence, is to be optimized, where in  $m \leq n$  and  $m < M$ ,
- an algorithm for generation of one or more further test sequences from the first test sequence by replacing at one or more of the m optimization positions a

- codon of the first test sequence by another codon which expresses the same amino acid,
- an algorithm for assessment of each of the test sequences with a quality function and for ascertaining the test sequence which is optimal in relation to the quality function,
- an algorithm for specification of p codons of the optimal test sequence which are located at one of the in optimization positions, as result codons which form the codons of the optimized nucleotide sequence at the positions which correspond to the positions of said p codons in the test sequence, where p is a natural number and  $p \leq m$ ,
- an algorithm for iteration of the steps of generation of a plurality of test functions, of assessment of the test sequences and of specification of result codons, where in each iteration step the test sequence comprises the appropriate result codon at the positions which correspond to positions of specified result codons in the optimized nucleotide sequence, and the optimization positions are different from positions of result codons.

12. (Previously presented) The device as claimed in claim 11, characterized by an algorithm for carrying out the steps of a method as claimed in claim 1.

13. (Previously presented) The device as claimed in claim 11, characterized by a device for automatic synthesis of nucleotide sequences which is controlled by the computer in such a way that it synthesizes the optimized nucleotide sequence.

V Product?

14. (Currently Amended) A computer program comprising instructions encoded on a non-transitory computer readable medium, wherein the instructions are executable by a computer to which comprises program code which can be executed by a computer and which, when it is executed on a computer, causes the computer to carry out a method as claimed in claim 1.

15. (Currently Amended) The computer program as claimed in claim 14, where the program code instructions can, when it is executed on by a computer, cause a device for the automatic synthesis of nucleotide sequences to prepare the optimized nucleotide sequence.

16. (Cancelled)

17. (Withdrawn) A nucleic acid which includes a nucleotide sequence coding for a protein and which is obtainable by a method as claimed in claim 9.

18. (Withdrawn) The nucleic acid as claimed in claim 17, characterized in that the latter includes a nucleotide sequence which codes in a predefined organism for a protein, where said nucleotide sequence is not present in the naturally occurring genome of the organism.

19. (Withdrawn) The nucleic acid as claimed in claim 18, characterized in that the organism is selected from the following group:

- viruses, especially vaccinia viruses,
- prokaryotes, especially Escherichia, Caulobacter crescentus, *Bacillus subtilis*, *Mycobacterium spec.*,
- yeasts, especially *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia angusta*,
- insects, especially *Sphingoptera frugiperda*, *Drosophila spec.*,
- mammals, especially *Homo sapiens*, *Macaca mulata*, *Mus*

U.S. Serial Number 10/539,208  
Page 7

musculus, Bos taurus, Capra hircus, Ovis aries, Oryctolagus cuniculus, Rattus norvegicus, Chinese hamster ovary,

- monocotyledonous plants, especially Oryza sativa, Zea mays, Triticum aestivum,
- dicotyledonous plants, especially Glycin max, Gossypium hirsutum, Nicotiana tabacum, Arabidopsis thaliana, Solanum tuberosum.

20. (Withdrawn – previously presented) The nucleic acid as claimed in claim 4-17, characterized in that the protein encoded by the nucleotide sequence is one of the following proteins and/or falls into one of the following protein classes:

- enzymes, especially polymerases, endonucleases,
- ligases, ligases, ligases, proteases, kinases, phosphatases, topoisornerases, cytokines, chemokines, transcription factors, oncogenes,
- proteins from thermophilic organisms, from cryophilic organisms, from halophilic organisms, from acidophilic organisms, from basophilic organisms, proteins with repetitive sequence elements, especially structural proteins, human antigens, especially tumor antigens, tumor markers, autoimmune antigens, diagnostic markers,
- viral antigens, especially from HAV, HBV, HCV, HIV, SIV, H<sup>y</sup>, HPV, rinoviruses, influenza viruses, herpesviruses, pellenr-uses, polyoma viruses, hendra virus, dengue virus, AAV, adenoviruses, HTLV, RSV,

U.S. Serial Number 10/539,208  
Page 8

- antigens of disease-causing parasites, e.g. protozoa, especially those causing malaria, leishmania, trypanosoma, toxoplasmas, amoeba,
- antigens of disease-causing bacteria or bacterial pathogens, especially of the genera Chlamydia, staphylococci, Klebsiella, Streptococcus, Salmonella, Listeria, Borrelia, Escherichia coli,
- antigens of organisms of safety level L4, especially Bacillus anthracis, Ebola virus, Marburg virus, poxviruses.

21. (Withdrawn) The nucleic acid as claimed in claim 18, characterized in that the quality function takes account at least of one the following criteria:

- GC content,
- codon usage of the predefined organism,
- exclusion of inverse complementary sequence identities of more than 20 nucleotides to the transcriptome of a predetermined organism,
- complete or substantial exclusion of homology regions of more than 100 base pairs to a predefined DNA sequence,
- complete or substantial exclusion of homology regions with a similarity of more than 90% to a predefined DNA sequence.

22. (Withdrawn) A vector comprising a nucleic acid as claimed in claim 17.

23. (Withdrawn) A cell comprising a vector as claimed in claim 22.

24. (Withdrawn) An organism comprising at least one cell as claimed in claim 23.

25. (Withdrawn) A nucleic acid, in particular as claimed in claim 9, comprising a nucleotide sequence which is selected from the group comprising: SEQ ID NO: 2, 4, 6,

U.S. Serial Number 10/539,208  
Page 9

8.

26. (Withdrawn) A vector comprising a nucleic acid as claimed in claim 25.
27. (Withdrawn) A cell comprising a vector as claimed in claim 26.
28. (Withdrawn) An organism comprising at least one cell as claimed in claim 27.
29. (Withdrawn) A cell comprising a nucleic acid as claim in claim 17.
30. (Withdrawn) A cell comprising a nucleic acid as claimed in claim 25.